INFOTEHNOLOOGIA
Universaalne koodimärgistik (UCS)

Information technology
Universal coded character set (UCS)
(ISO/IEC 10646:2020, identical +
ISO/IEC 10646:2020/Amd 1:2023, identical)

EVS • EESTI STANDARDIMIS- JA AKREDITEERIMISKESKUS
ESTONIAN CENTRE FOR STANDARDISATION AND ACCREDITATION

EESTI STANDARDI EESSÕNA	NATIONAL FOREWORD
See Eesti standard EVS-ISO/IEC 10646:2025 sisaldab rahvusvahelise standardi ISO/IEC 10646:2020 „Information technology — Universal coded character set (UCS)“ ning selle muudatuse ISO/IEC 10646:2020/Amd 1:2023 identset ingliskeelset teksti.	This Estonian Standard EVS-ISO/IEC 10646:2025 consists of the identical English text of the International Standard ISO/IEC 10646:2020 „Information technology — Universal coded character set (UCS)“ including its Amendment ISO/IEC 10646:2020/Amd 1:2023.
Ettepaneku rahvusvahelise standardi ümbertrüki meetodil ülevõtuks on esitanud EVS/TK 04 standardi avaldamist on korraldanud Eesti Standardimis- ja Akrediteerimiskeskus.	Proposal to adopt the International Standard by reprint method has been presented by EVS/TC 04, the Estonian Standard has been published by the Estonian Centre for Standardisation and Accreditation.
Standard EVS-ISO/IEC 10646:2025 on jõustunud sellekohase teate avaldamisega EVS Teatajas.	Standard EVS-ISO/IEC 10646:2025 has been endorsed with a notification published in the official bulletin of the Estonian Centre for Standardisation and Accreditation.
Standard on kättesaadav Eesti Standardimis- ja Akrediteerimiskeskusest.	This standard is available from the Estonian Centre for Standardisation and Accreditation.

Käsitlusala

See dokument

— täpsustab UCSi struktuuri;

— määratleb UCSi kohta kasutatavad terminid;

— kirjeldab UCSi koodiruumi üldstruktuuri;

— kirjeldab määratletud UCSi tasandeid: mitmekeelne põhitasand (*Basic Multilingual Plane*, BMP), mitmekeelne lisatasand (*Supplementary Multilingual Plane*, SMP), ideograafiline lisatasand (*Supplementary Ideographic Plane*, SIP), tertsiaarne lisatasand (*Tertiary Ideographic Plane*, TIP) ja eriotstarbeline lisatasand (*Supplementary Special-purpose Plane*, SSP);

— määratleb kirjamärkide kogumi, mida kasutatakse ülemaailmselt skriptides ja loomulike keelte kirjapildis;

— täpsustab kirjamärkide ja vormingumärkide nimesid BMP, SMP, SIP, TIP, SSP ning nende kodeeritud esituste jaoks UCS-koodiruumis;

— täpsustab juhtmärkide ja privaatmärkide kodeeritud esitust;

— täpsustab kolme UCSi kodeerimisvormi: UTF-8, UTF-16 ja UTF-32;

— täpsustab seitset UCSi kodeerimisskeemi: UTF-8, UTF-16, UTF-16BE, UTF-16LE, UTF-32, UTF-32BE ja UTF-32LE;

— täpsustab selle koodimärgistiku tulevaste lisandite haldust.

MÄRKUS Dokument ei täpsusta, kuidas selgitada välja kirjeldatud märkide sobivust identifikaatoriteks programmeerimiskeeltes, kuid seda küsimust käsitletakse dokumendis viidatud allikas, vt lisa U.

Tagasisidet standardi sisu kohta on võimalik edastada, kasutades EVS-i veebilehel asuvat tagasiside vormi või saates e-kirja meiliaadressile standardiosakond@evs.ee.

ICS 35.040.10

[Blank page]

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see http://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology,* Subcommittee SC 2, *Coded character sets.*

This sixth edition of ISO/IEC 10646 cancels and replaces the fifth edition (ISO/IEC 10646:2017), which has been technically revised. It also incorporates ISO/IEC 10646:2017/Amd 1:2019 and ISO/IEC 10646:2017/Amd 2:2019.

This edition includes the following significant changes with respect to the previous edition:

- new scripts covered: Chorasmian, Dives Akuru, Dogra, Elymaic, Gunjala Gondi, Hanifi Rohingya, Khitan Small Script, Makasar, Medefaidrin, Nandinagari, Nyiakeng Puachue Hmong, Old Sogdian, Sogdian, Yezidi, Wancho;
- existing scripts significantly extended: Georgian, CJK Unified Ideographs (Extension G);
- new symbol sets: Chess Symbols, Symbols for Legacy Computing;
- new set of Emoji symbols.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document specifies the Universal Coded Character Set (UCS). It is applicable to the representation, transmission, interchange, processing, storage, input and presentation of the written form of the languages of the world as well as additional symbols.

By defining a consistent way of encoding multilingual text it enables the exchange of data internationally. The information technology industry gains data stability, greater global interoperability and data interchange. This International Standard has been widely adopted in new Internet protocols and implemented in modern operating systems and computer languages. This edition covers over 130 000 characters from the world's scripts.

The UCS is an encoding system different from that specified in ISO/IEC 2022. The method to designate UCS from ISO/IEC 2022 is specified in 13.2.

A graphic character will be assigned only one code point in the standard, located either in the BMP or in one of the supplementary planes.

Information technology — Universal Coded Character Set (UCS)

1 Scope

This document

- specifies the architecture of the UCS;
- defines terms used for the UCS;
- describes the general structure of the UCS codespace;
- specifies the assigned planes of the UCS: the Basic Multilingual Plane (BMP) of the UCS, the Supplementary Multilingual Plane (SMP), the Supplementary Ideographic Plane (SIP), the Tertiary Ideographic Plane (TIP), and the Supplementary Special-purpose Plane (SSP);
- defines a set of graphic characters used in scripts and the written form of languages on a world-wide scale;
- specifies the names for the graphic characters and format characters of the BMP, SMP, SIP, TIP, SSP and their coded representations within the UCS codespace;
- specifies the coded representations for control characters and private use characters;
- specifies three encoding forms of the UCS: UTF-8, UTF-16, and UTF-32;
- specifies seven encoding schemes of the UCS: UTF-8, UTF-16, UTF-16BE, UTF-16LE, UTF-32, UTF-32BE, and UTF-32LE;
- specifies the management of future additions to this coded character set.

NOTE The determination of suitability of these characters for use as identifiers in programming languages is not specified by this document but can be found in an external reference. See Annex U.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 2022 *Information technology — Character code structure and extension techniques*

ISO/IEC 6429 *Information technology — Control functions for coded character sets*

Unicode Standard Annex, UAX #9, *The Unicode Bidirectional Algorithm:*
http://www.unicode.org/reports/tr9/tr9-42.html

Unicode Standard Annex, UAX #15, *Unicode Normalization Forms:*
http://www.unicode.org/reports/tr15/tr15-50.html

Unicode Technical Standard, *UTS #37, Ideographic Variation Database:*
http://www.unicode.org/reports/tr37/tr37-12.html

Unicode Standard Version 13.0*, Chapter 4, Character Properties*
http://www.unicode.org/versions/Unicode13.0.0/ch04.pdf
Section 4.3, Combining Classes – Normative
Section 4.5, General Category – Normative
Section 4.7, Bidi Mirrored – Normative

Unicode Standard Version 12.1, *Age Property:*
https://www.unicode.org/Public/13.0.0/ucd/DerivedAge.txt

NOTE Parts of this document which use machine-readable format are available as electronic data attachments. See Clause 5.